A mathematical model to correlate the importance of gene specific mutations and tumor development

Kevin Flores (Arizona State University), Yi Jiang(LANL)

Understanding the correlation of gene specific mutations and tumor development has important implications in cancer therapy. Recent empirical data have elucidated the candidate cancer genes responsible for carcinogenesis through mutation and expression analysis [1]. This work has revealed the heterogeneities in genotype that encode cancers of the same malignancy grade, providing evidence for the existence of multiple mutational paths that a population of cancer cells can take to manifest itself as a disease.

The cell genotypes that are present in a tumor affect the malignancy grade through their effect on the phenotypes of individual cells that the tumor is comprised of. We use a graph theoretical approach to connect the gene mutation data to cell phenotype; we postulate that the detected gene mutations correspond to inactivated or overexpressed genes that encode proteins that affect the malignant cell phenotype. We have constructed a gene regulatory network from the KEGG pathway database [2], (Figure 1). This network includes most accurately and completely the relevant pathways that contain the known cancer genes, which in turn encode distinct cell phenotypes. We are analyzing the network to predict the sensitivity of cell signaling pathways that control cell growth and death to alterations caused by gene mutations.

We have chosen specific proteins in the network whose presence determine the phenotype of a cell containing a network of the interacting proteins included in this model. The phenotypic states and their respective proteins that are known experimentally to have

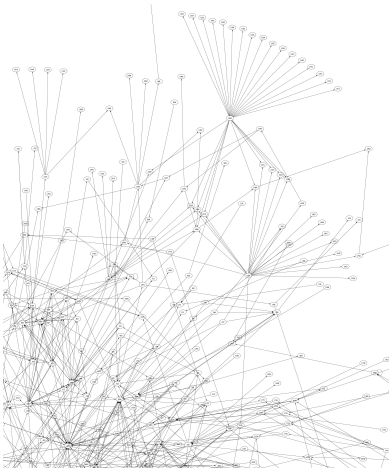

Figure 1: A piece of the signal transduction network of 807 nodes and 1593 edges.

a direct affect on cell phenotype are: Proliferation(CDK1,CDK2,CDK4,CDK6), Apoptosis(CASP3,CASP7), Survival(BCL2,BC-xl), and Anti-Survival(BAD,BAX). The prevalence of gene mutations showed no correlation to simple measures of their equivalent representations in the network. The betweenness-centrality measure of genes in the network show no correlation to the percent of tumors from the Sjoblom et al experiments that contain the mutations; this shows that the centrality of genes in the network has no direct affect on the malignant cell phenotype. We also calculated the amount of paths from any of the candidate genes to specific proliferation effector proteins(CDK's) and their totals; this also showed no correlation to an oncogenic effect implied by the percent of tumors containing the mutated candidate genes. This suggests that measures of the static network are insufficient to explain the importance of the mutated genes implied by their observed unusually high mutation frequency, thus a dynamical model is required. Because of the lack of necessary reaction rate data to model any of the interactions, we turn to a network boolean dynamics model in which the state of proteins, represented by nodes in the network, are on or off and are updated in

A mathematical model to correlate the importance of gene specific mutations and tumor development

time using functions depending on the network connections.

When all the nodes are updated simultaneously at each time step we find that the phenotypic output resulting from the deterministic network dynamics are insensitive to the candidate gene mutations. We represent the deletion mutations in the simulations by deleting the mutated genes' node and incident edges; the mutations causing overexpression of a gene are represented by fixing the state of that gene on throughout the simulation. With nonsimultaneous updating we find that the state space of the dynamics becomes too large to sample using random initial conditions. We employ the Wang-Landau monte carlo algorithm to determine the proportions of inital conditions that have attractors whose protein profiles(on or off) are classified into distinct phenotypes [3]. We consider 3 distinct cell phenotypes: Proliferation, Apoptosis, and Survival (Figure 2). With this algorithm we estimate

Proliferation	1	0	0	0	1	0	0	1	0	1	1	0	1	1	1	0	
Apoptosis	0	1	0	0	1	1	0	0	1	0	1	1	0	1	1	0	
Survival	0	0	1	0	0	1	1	1	0	0	1	1	1	0	1	0	
Anti-Survival	0	0	0	1	0	0	1	0	1	1	0	1	1	1	1	0	
Phenotype	P	A	S	S	A	S	S	A	S	S	P	A	P	A	A	S	

Figure 2: The Phenotypes(P=Proliferating, A=Apoptosis, S=Survival) are determined by the state(on = 1, off = 0) of the families of effector proteins (shown in 2) in the network. For example, if all of the proliferation proteins(cdks) are on then the table has a 1 in the proliferation row, and a 0 if they are not all on; the state of the 4 families in each column determines a phenotype.

that the entire state space, whose size is on the order of 2^{800}, can be sampled with less than the equivalent of 30,000 random initial conditions when we want to determine the proportions of the state space that lead to any of the predefined phenotypes. With this type of sampling we can determine whether changes in the network caused by mutations lead to altered proportions of the total amount of possible network states whose progression will end in the distinct phenotypes; thus we can correlate mutations to the increased malignancy of a cell by measuring whether the proportions of states ending in proliferation and survival states increases and those ending in apoptosis states decrease.

Acknowledgements

This work was carried out under the auspices of the National Nuclear Security Administratioin of the U.S. Department of Energy at Los Alamos National Laboratory under contract No. DE-AC52-06NA25396.

References

[1] Sjoblom T. et al. Science 13 October 2006 314: 268-274.

[2] Kanehisa M., Goto S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27-30 (2000).

[3] Wang F., Landau D.P.; Physical Review Letters 5 March 2001 86,10: 2050-2053